RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/988,384B
Input Set: A:\600536US1.txt
Output Set: N:\CRF3\06112002\1988384B.raw
CANT: Sherman, D.H.

```
4 <110> APPLICANT: Sherman, D.H.
        Liu, H.
9 <120> TITLE OF INVENTION: DNA encoding methymycin and pikromycin
5
11 <130> FILE REFERENCE: 600.536US1
13 <140> CURRENT APPLICATION NUMBER: US 09/988,384B
14 <141> CURRENT FILING DATE: 2001-11-19
16 <150> PRIOR APPLICATION NUMBER: PCT/US99/14398
17 <151> PRIOR FILING DATE: 1999-06-25
19 <150> PRIOR APPLICATION NUMBER: US 09/105,537
20 <151> PRIOR FILING DATE: 1998-06-26
 22 <160> NUMBER OF SEQ ID NOS: 53
 24 <210> SEQ ID NO: 1
 25 <211> LENGTH: 15872
 27 <213> ORGANISM: Streptomyces venezuelae
 26 <212> TYPE: DNA
  30 ttaattaagg aggaccatca tgaacgaggc catcgccgtc gtcggcatgt cctgccgcct
                                                                             60
                                                                            120
  31 gccgaaggcc tcgaacccgg ccgccttctg ggagctgctg cggaacgggg agagcgccgt
 29 <400> SEQUENCE: 1
                                                                            180
  32 caccgacgtg ccctccggcc ggtggacgtc ggtgctcggg ggagcggacg ccgaggagcc
  33 ggcggagtcc ggtgtccgcc ggggcggctt cctcgactcc ctcgacctct tcgacgcggc
                                                                            240
                                                                            300
  34 ettettegga atetegeece gtgaggeege egecatggae eegeageage gaetggteet
                                                                             360
   35 cgaactcgcc tgggaggcgc tggaggacgc cggaatcgtc cccggcaccc tcgccggaag
   36 ecgeacegee gtettegteg geacectgeg ggacgactae acgageetee tetaceagea
                                                                             420
   37 cggcgagcag gccatcaccc agcacaccat ggcgggcgtg aaccggggcg tcatcgccaa
                                                                             480
   38 cegestetes taccacetes geetscages ecesageete accestegaces cegescages
                                                                             540
   39 gtcctcgctc gtcgccgtgc acctggcctg cgagtccctg cgcgccgggg agtccacgac
                                                                             600
   40 ggcgctcgtc gccggcgtga acctcaacat cctcgcggag agcgccgtga cggaggagcg
                                                                             660
                                                                              720
    41 cttcggtgga ctctccccgg acggcaccgc ctacaccttc gacgcgcggg ccaacggatt
    42 cgtccggggc gagggggg gagtcgtcgt actcaagccg ctctcccgcg ccctcgccga
                                                                              780
    43 eggegacegt gtecaeggeg teateegege cagegeegte aacaaegaeg gageeaecee
                                                                              840
    44 gggtctcacc gtgcccagca gggccgccca ggagaaggtg ctgcgcgagg cgtaccggaa
                                                                              900
    45 ggcggccetg gacccgtccg ccgtccagta cgtcgaactc cacggcaccg gaaccccgt
                                                                              960
    46 cggcgacccc atcgaggccg ccgcgctcgg cgccgtcctc ggctcggcgc gccccgcgga
                                                                              1020
    47 cgaacccctg ctcgtcggct cggccaagac gaacgtcggg cacctcgaag gcgccgcgg
                                                                              1080
     48 catcgtcggc ctcatcaaga cgctcctcgc gctcggccgg cgccggatcc cggcgagcct
                                                                              1140
     49 caactteegt aegeeceace eggacateee getegacace etegggeteg aegtgeecga
                                                                              1200
     50 eggeetgegg gagtggeege acceggaceg egaacteete geeggegtea getegttegg
                                                                              1260
     51 catgggeggc accaacgccc acgtcgtcct cagcgaaggc cccgcccagg gcggggagca
                                                                              1320
     52 gcccggcatc gatgaggaga cccccgtcga cagcggggcc gcactgccct tcgtcgtcac
                                                                              1380
                                                                              1440
      53 cggccgcggc ggcgaggccc tgcgcgccca ggcccggcgc ctgcacgagg ccgtcgaagc
                                                                               1500
      54 ggacceggag etegegeceg eegcactege eeggtegetg gteaccacee gtacggtett
```

RAW SEQUENCE LISTING PATENT APPLICATION:

Input Set : A:\600536US1.txt

55 cacgcaccgg teggtegtee tegeceegga cegegeege etecteg 53:12
56 cetegeegee gggacgeeg egeceggegt ggteaccqqc acceenges
57 cgcqqtectq 57 egecgteetg ticageggee agggtgeeca acgtaeggge atgggeatgg as 58 egeccaecce geettegega eggeettega egeegtegee geegaactgg acce 59 cgaccggcc ctcgccgaac tcgtcgcggc gggcgacacc ctcgaccgca ccgtccolson 59 cgaccggccc ctcgccgaac tcgtcgcgc gggcgacacc ctcgaccgca ccgtccator ctcgaccgca cctggggcg20
60 acagcccgcg ctcttcgccg tggaggtcgc cctccaccgc acgtcgcgca acgtcgccac 61 cacgcccgac ctgctcgccg gccactccgt cggcgagatc agcgccgcc acgtcgccgg 62 ggtcctgtcg ctgcgcgacg ccgcccgcct cgtcgcggcg cgcggccgcc tcatgcaggc 63 gctccccgag ggcggcgcga tggtcgcggt cgaggcgagc gaggaggaag tgcttccgca 204 64 cctcgcggga cgcgagcggg agctctccct cgcggccgtg aacggccccc gcgcggtcgt 2100 65 cctcgcgggc gccgagcgcg ccgtcctcga cgtcgccgag ctgctgcgcg aacagggccg 2160 66 ccggacgaag cggctcagcg tetcgcacgc cttccactcg ccgctcatgg agccgatgct 2220 67 cgacgaette cgccgggtcg tegaagaget ggacttecag gageeeegeg tegacgtegt 2280 68 gtccacggtg acgggcctgc ctgtcacagc gggccaatgg accgatcccg agtactgggt 2340 69 ggaccaggtc cgcaggcccg tacgcttcct cgacgccgta cgcaccctgg aggaatcggg 2400 70 cgccgacacc ttcctggagc tcggtcccga cggggtctgc tccgcgatgg cggcggactc 2460 2520 71 cgtacgcgac caggaggccg ccacggcggt ctccgccctg cgcaagggcc gcccggagcc 72 ccagtcgctg ctcgccgcac tcaccaccgt cttcgtccgg ggccacgacg tcgactggac 2580 2640 73 cgccgcgcac gggagcaccg gcacggtcag ggtgcccctg ccgacctacg ccttccagcg 74 cgaacgccac tggttcgacg gcgccgcgcg aacggcggcg ccgctcacgg cgggccgatc 2700 75 gggcaccggt gcgggcaccg gcccggccgc gggtgtgacg tcgggcgagg gcgagggcga 2760 76 gggcgagggc gcgggtgcgg gtggcggtga tcggccggct cgccacgaga cgaccgagcg 2820 2880 77 cgtgcgcgca cacgtcgccg ccgtcctcga gtacgacgac ccgacccgcg tcgaactcgg 2940 78 cctcaccttc aaggagctgg gcttcgactc cctcatgtcc gtcgagctgc ggaacgcgct 79 cgtcgacgac acgggactgc gcctgcccag cggactgctc ttcgaccacc cgacgccgcg 3000 80 cgccctcgcc gcccacctgg gcgacctgct caccggcggc agcggcgaga ccggatcggc 3060 81 cgacgggata ccgcccgcga ccccggcgga caccaccgcc gagcccatcg cgatcatcgg 3120 82 catggcctgc cgctaccccg geggcgtcac ctcccccgag gacctgtggc ggctcgtcgc 3180 83 cgaggggcgc gacgccgtct cggggctgcc caccgaccgc ggctgggacg aggacctctt 3240 84 cgacgccgac cccgaccgca gcggcaagag ctcggtccgc gagggcggat tcctgcacga 3300 3360 85 cgccgccctg ttcgacgccg gcttcttcgg gatatcgccc cgcgaggccc tcggcatgga 86 cccgcagcag cggctgctcc tggagacggc atgggaggcc gtggagcgcg cagggctcga 3420 87 ccccgaaggc ctcaagggca gccggacggc cgtcttcgtc ggcgccaccg ccctggacta 3480 88 cggcccgcgc atgcacgacg gcgccgaggg cgtcgagggc cacctcctga ccgggaccac 3540 89 gcccagcgtg atgtcgggcc gcatcgccta ccagctcggc ctcaccggtc ctgcggtcac 3600 90 cgtcgacacg gcctgctcgt cetcgctcgt cgcgctgcac ctggccgtcc gttcgctgcg 3660 91 gcagggcgag tcgagcctcg cgctcgccgg cggagcgacc gtcatgtcga caccgggcat 3720 92 gttcgtcgag ttctcgcggc agcgcggcct cgccgccgac ggccgctcca aggccttctc 3780 93 cgactccgcc gacggcacct cctgggccga gggcgtcggc ctcctcgtcg tcgagcggct 3840 94 ctcggacgcc gagcgcaacg gccaccccgt gctcgccgtg atccggggca gcgcggtcaa 3900 95 ccaggacggc gcctccaacg ggctcaccgc ccccaacggc ccgtcccagc agcgcgtcat 3960 96 ccgacaggcc ctggccgacg ccgggctcac cccggccgac gtcgacgccg tcgaggcgca 4020 97 cggtacgggt acceggeteg gegaccecat cgaggeegag gegateeteg geacetacgg 4080 4140 4200 99 cgcgcaggcc gccgcgggcg tgggcgggct catcaagatg gtcctcgcga tgcgccacgg 100 cgtcctgccc aggacgctcc acgtggaccg gcccaccacc cgcgtcgact gggaggccgg 4260 4320 102 ggcgatetec teetteggea teageggeae caaegeecae ategtggteg aacaggeece 4380 103 ggaagccggg gaggcggcgg tcaccaccac cgccccggaa gcaggggaag ccggggaagc 4440

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/988,38

Input Set : A:\600536US1.txt

Input Set : A:\600536US1.txt Output Set: N:\CPEN 0.5
Output Set: N:\CRF3\06112002\I988384B.ra\1/2002
104 ggggg
104 ggcggacacc accgccacca cgacgccggc cgcggtcggc gtccccgaac c.  105 cccgtcgtg gtctccgcgc gggacgccgc cgcctgcgc gtccccgaac c.  106 gaccttcctc gacggccgac cggacgtcg
107 gacettecte gacggegge egectgege egecetgege egectgege egetge eg
106 cgggctcgac gcctcggc aggccgcct caccacggc accact cgctga
107 ccgtaccgcc ttcgagcaca cggacgtcac cgtcgccgac ccaggccg ttcgacgac ttcgagcaca aggccgcct caccaccgcc accagggacg agctgctc4500 cgaggaactg cgcacggcc tcctgttcac cggccaggga cgcacggcc acccgtgt cgcaggga gcgcaggga agccacggga cgcacggcc acccgtgt cgcaggga gcgcaggga acccgtgtt cgccaggga gcgcagcgcg tcgcatggg 112 cctcaccgga taccacgac ggccgctggg gagacacgg tgaacacgg tacacgg tgaacacgg tacacgg tacacg
cgaggaactg cgcgccgcgc tcctgttcac cggccaggag ctggtcaccg gcgaaccgg60
111 cctcgaccgt cacct caccatgtt cgccgagag degcagegeg tcgccatgg
110 cgaggaactg cgcacggcct tcctgttcac cggccagggc ctggtcaccg gcgaaccggc60 111 cctcgaccgt cacctcgacc ggccgctgcg ggagatcgtc ctcgacaccg tcgccatggg 112 cctcaccgcg tacacccagc cgcctctt ggagaaccgg ggagatcgtc gccgccgggg aggagatcg
111 cctcgaccgt cacctcgacc accccgtgtt cgccgccgcc ctcgacaccg tcgccatggg 112 cctcaccgcg tacacccagc cgccctctt cgccttcgac gcgcaccggg aggagatcgtc ccgacaccg tgtacgcggc tgtacgcggc 113 cgaacaccac ggcctcgtc ccgacctctt cgccttcgag gtggcgctgt tcgccatggg 114 cgcgcacgtc gcgactgtc ccgacctgct caccggccac tccgtcggg aggagatcggc 115 ccggctcatg cagtcggcc cgacctgct caccggccac tccgtcggg aggagatcgcc 116 cgaggtcgtc cagtcggcc gcgagggggg cgacgcgca cgtctcgtca cgcaccgc 5040
114 cgcgcacgtc gcgcctgtc ccgacctgct cacagagas gtggcgctgt tcggcstart
115 ccggctcatg cagter tetecetega cgaggera tecgteggeg agategor 4980
114 cgcgcacgtc gccggtgtcc ccgacctgct caccggccac tccgtcggg gtggcgctgt tccgcctct 4980 115 ccggctcatg cagtcggcc gcgagggcgg cgcgatgatc gcgccgcac gagtcgcga gggcggggg cgcgatgatc gcgcgcgacgc gggcggggg cgcgatgatc gcgcgcgggggggggg
117 acccaccacc structure agggetacga agggetacga gggggggggggggggggggggggggggggggggg
118 ggcgggacga ggcggacga ggggggggggggggggg
119 catggargae ggccggcga cccgcagggt gaggcgcgc gaggagatcc ggccaacgg 5220
120 gcggst- 9tcctcgacg agttosts 9cgcgtcagc cacgcgttg 5280
121 gccgccd gcgtctcca cggtcacca ggccgag ggcctgacct dctccccgca 5340
122 cotago de Egggtogaco agatogaga eggetogaco acatogagagago 5400
123 actores daggacgega cogtostant geologicge ttcctggacg
124 categorate yaggeteteg cacegorate cyagategge cecagorate 5520
125 cgcgcgcgggggggggggggggggggggggggggggg
126 ggacet Cacggcgcac ccttggagacc ctcgccgagacc ctcgccgagac
127 granges acgtacgect technology gyegtegtte ttechnology 5700
128 ggtggggg t tggaggggg ttgaggggg 5760
129 gctggacggcq tcctggtgac 5820
130 caccata Cacatagtea acadegacata tegetal tecetagaca 5880
131 gccgot and gcgggggacc acgtcgggggggacc acgtcgggggggggg
132 cgagtan Ctgcccgagc gggggggggggggggggggggggggggggg
133 transport georgian contract contrac
134 ggggaegg ccccgggggt ggacggg gtacagcacc cccgactgg 6120
135 ggcggaccacc ccggcacac tylctccggc gtactcggc 6180
136 ataggat gccacaccc togacgas cagggacgt tcagggggt tcaggggct 6240
137 cgagata geottecagg geotgagag etaggagag etaggagagagagagagagagagagagagagagagagaga
138 gata-1 339 CtgCcgcgg cgcagagagagagagagagagagagagagagaga
139 tgaccarates gagagagagagagagagagagagagagagagagagaga
140 cctoss decagacgaga ccgacgagagagagagagagagagagagag
141 ggtcacgctc ggagggcca ccgcgctccg cgtacac ttcgcgtggg cggggctcg 6540
139 tgacgcegat accgacgga cegaceggat cetteggeg tacacecegge 6420 140 cetecacgec gaaggggea cegaceggat ceggetgeeg tacacecegge 6540 141 ggtcacgete egeetcaceg acaccacegg tgegeggeggggggggggggggggggggggggggg
143 gttcacggtg accggctggg accggctggg gccaccgtgg agtcact 6660
144 ggtcgaggaa at a state of the state of th
142 cctgcgcgcg gtggcgaagg accgctcgg tgcgcccgtg gcacccaccgg gcccggacac acaccaccgg tgcgcccgtg gccaccgtgg agtccctgac agaccggcac accgaccgcc ggggggcgcc gggggggacaccaccgcg gggggggacaccaccgcg gggggggacaccaccgcg gggcggggacaccaccgcg gggggggacaccaccggcgacaccaccgg gggcggggacaccaccggcaccaccggacaccaccggacaccac
145 cgtcgtcetc cgggccgacc tcgccggcct cggcgacctc gcagggcgcg gagccgtgga 6840 146 agccgtccgg cgcaccctcg ccatcgtcca ggagttcctg tcgaggcgcg cggcgcgga 6900 147 ctcgcgactg gtgtgcgtca ccagggggg ggagttcctg tccgaggcgc gcttaggaggg 6960
147 ctcgcgacts of coatcgtcca catcgtcca catcgtca
148 ctccttgges gtgtgcgtca ccaggggggg
146 agccgtccgg cgcacctcg gctggacget cgacggggac cgacggggacgggacggggacggggacggggacggggacggggacggggacgggggg
150 gatccgcgg acceptage tgggtgaagg acceptage georgeteg aggacgteac 7080
149 acgectgtte etcetggace tetggggeet egteegetee geeragteeg agaacgteae 7020 150 gateegegeg geeraggeeg gegaaggegag egegaeggag eggaeggag eggaegga
152 ccgtttcggg care ccgccgct the control of the con
150 gatecgegeg gecaeggeg teggegegegegegegegegegegegegegegeg
7260 7260 7260 7260 7260 7260 7260 7260
7320 accygaggce teggageget
7380

RAW SEQUENCE LISTING DATE: 06/11/2002 PATENT APPLICATION: US/09/988,384B TIME: 16:53:12

Input Set : A:\600536US1.txt

Output Set: N:\CRF3\06112002\I988384B.raw

153	cctcgcccgc	cacctcgtgg	agcgtcacgg	ggtgcgccgg	ctgctgctgg	tgagccgccg	7440
154	cggggccgac	gcccccggcg	cggccgacct	gggcgaggac	ctcgcgggcc	tcggcgcgga	7500
					ctggcgcggg		7560
					gcgggagtcg		. 7620
157	gacggtggag	gcgctcacac	cggaacggct	ggacgcggta	ctgcgcccga	aggtcgacgc	7680
					gacgccttcg		7740
					tacgcggcgg		7800
					gccgccacgt		7860
					ggcgccgccg		7920
					ggcctcgcgc		7980
					ctccgtccca		8040
					gtccgggcgc		8100
					ggctggctga		8160
					ctcgtgacgg		8220
					gagcggtcct		8280
					ctgaacgccg		8340
					gccgcgctcg		8400
					accgcgaccg		8460
					ccgatcgcga		8520
					ctgtggcggc		8580
					tgggacctgg		8640
					gagggcgggt		8700
					cgcgaggccc		8760
					ctggaacgcg		8820
					ggcgccgcgc		8880
					tatctgctga		8940
					ctcgaaggac		9000
					ctggcggtac		9060
					gtgatggccg		9120
					ggccgctgca		9180
					gtgctcgtcc		9240
					gtcaccggca		9300
					ccagcccagc		9360
					gtggacgcgg		9420
					gcgctgctcg		9480
					ctgaagtcca		9540
					ctccaggcgc		9600
					catgccgact		9660
					accggccggc		9720
					gtcgtcctcg		9780
					ggctcccgcg		9840
					gagccggccc		9900
195	ggcccggcgg	ctccgcgacc	acctctcccg	cacccccggg	gcccgcccgc	gtgacatcgc	9960
196	cttctccctc	gccgccacgc	gcgcagcctt	tgaccaccgc	gccgtgctga	teggetegga	10020
19/	cggggccgaa	ctcgccgccg	ccctggacgc	gttggccgaa	ggacgcgacg	gtccggcggt	10080
198	ggtgcgcgga	gtccgcgacc	gggacggcag	gatggccttc	ctcttcaccg	ggcagggcag	10140
T33	ccagcgcgcc	gggatggccc	acgacctgca	tgccgcccat	accttcttcg	cgtccgccct	10200
200	cgacgaggtg	acggaccgtc	tcgacccgct	geteggeegg	ccgctcggcg	cgctgctgga	10260
∠U1	egecegacee	ggctcgcccg	aagcggcact	cctggaccgg	accgagtaca	cccagccggc	10320

RAW SEQUENCE LISTING DATE: 06/11/2002 PATENT APPLICATION: US/09/988,384B TIME: 16:53:12

Input Set : A:\600536US1.txt

Output Set: N:\CRF3\06112002\1988384B.raw

202	gctcttcgcc	gtcgaggtgg	cgctccaccg	gctgctggag	cactggggga	tgcgccccga	10380
203	cctgctgctg	gggcactcgg	tgggcgaact	ggcggccgcc	cacgtcgcgg	gtgtgctcga	10440
					ctgatgcagc		10500
					gtccgcgcac		10560
					cggtcggtgg		10620
207	cgcggaggaa	gcggtggccg	aggcggcggc	gcagctcgcc	ggacgaggcc	gccgcaccag	10680
					gacggcatgc		10740
					ctgacggtcg		10800
					gactactggg		10860
					caccgcctcg		10920
					gcagaggagt		10980
				_	accgcgccgc	_	11040
					cggggcgccc		11100
					cgcgtcgaac		11160
					gcggcggaca		11220
					ggcggggccc		11280
					gagctgtccg		11340
					cccgtggacg		11400
					ggacctgaca		11460
					gaggaggatc		11520
					tccctgatgc		11580
					gccgcggtcg		11640
					ctggggcggg		11700
					gcctcgcccg		11760
					cagatcgccg		11820
					ggcggcggcc		11880
					ggcaccggcg		11940
					ctggcgctca		12000
					ctcctcctcc		12060
					ctcgcccggg		12120
					gcgggcgtac		12180
					tacgcgggca		12240
					ggactcgacg		12300
					ggtgcctacg		12360
					ggactccccg		12420
					gaggcgggcg		12480
					ctccgggcga		12540
					gactgggagc		12600
					ctcggcacgc		12660
					gcggcggacc		12720
					cgagcccgga		12780
					cggttccccg		12840
					tcggccacgg		12900
					ctcgccgccc		12960
					gggaccggtg		13020
					gtccgcaccg		13080
248	ggaaatggcc	gaactgctcg	ccqccqcttc	ccqqttccqc	cccgccttcg	qqacqqcqqa	13140
					gaggacaccg		13200
					gtggagttca		13260
		, , ,,,		,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,		, ,	

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/988,384B

DATE: 06/11/2002

TIME: 16:53:13

Input Set : A:\600536US1.txt

Output Set: N:\CRF3\06112002\I988384B.raw

L:4646 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order! L:4650 M:258 W: Mandatory Feature missing, <220> not found for SEQ ID#:25